**Lab Virtual Screen 3: Setting up the protein file for PTP1b in GOLD/HERMES**

**Objective**

The purpose of this lab is to screen a library of compounds for their ability to bind the active site of PTP1b enzyme as predicted by the GOLD docking program. The library will contain predominantly novel compounds that have not been tested against PTP1b in the wet lab. There will also be 5 known inhibitors of PTP1b taken from the BindingDB.com database as well as 5 compounds that have been selected in VDS from the Fall.

**LIBRARY SELECTION**

In a drug discovery project, it is advantageous to find new and unique scaffolds that are good inhibitors. A scaffold is basically the ‘backbone’ of the ligand and constitutes its core physic-chemical characteristics, in particular its steric properties. Ideally, new scaffolds will have different pharmacological effects than existing drugs and may, therefore, offer alternative therapies. Consequently, the Chembridge diversity set of ligands was put together to satisfy a wide range of scaffolds. Ultimately, once some hits are found, the researcher can then optimize a scaffold by adding, changing or deleting moieties and atoms on the outside.

The Chembridge Diveristy set consists of approximately 50,000 ligands that are derived from many different scaffolds. This library has been divided up into blocks of around 5,000 for each of you to screen against our target: the PTP1b enzyme. In additions to these untested compounds, it is also important to include some known active compounds in a virtual screen. Five inhibitors from that have been shown empirically to inhibit PTP1b have been added to the libraries. The 5 were taken from the BindingDB database (<http://www.bindingdb.org/bind/index.jsp>) amongst 1,760 listed inhibitors of PTP1b. Lastly, 4 compounds that have been tested by VDS researchers in 2010 and 2009 have been added to the library.

To relate this virtual screening lab to the wet lab that will be done in parallel, the compounds tested in the Enzyme Inhibition Assay consist of the 4 prior VDS compounds, plus one unknown from the Chembridge Diversity Set, plus one positive control from another source (orthovanadate). The BindingDB compounds are not included in the wet lab – but rather serve only as a positive control for virtual screening.

**Select one of the 10 libraries to screen**.

**CB5k\_1.sdf CB5k\_2.sdf CB5k\_3.sdf CB5k\_4.sdf CB5k\_5.sdf**

**CB5k\_6.sdf CB5k\_7.sdf CB5k\_8.sdf CB5k\_9.sdf CB5k\_10.sdf**

Use the **countsdf.pl** script in the LabVS3\_Library folder to find out how many ligands are present for the one you selected.

To run this script (sort of like a mini-program) type:

**$perl countsdf.pl**

These are the 4 prior VDS compounds which have been placed into each CB5k library already.

**5380289** – Chembridge **HTS09305SC** - Maybridge

**5753084** – Chembridge **SPB03365SC** - Maybridge

These are the 5 known inhibitors from BindingDB.com (already in each CB5k library).

**BindingDB\_50243240 BindingDB\_50131107 BindingDB\_50170986**

**BindingDB\_50131106 BindingDB\_50228029**

Lastly, here is the one unknown (untested) compound from Chembridge Diversity Set (already in each CB5k library)

**5648649** - Chembridge

**VIRTUAL SCREENING**

* 1st Run
  + Screen ~5,000 ligands at 0.1 autoscale on 6 processors
  + Save 10% = ~ 500
* 2nd Run
  + Screen the 500 ligands at 1 autoscale on 6 processors
  + Save 10% = 50
* Analyze Bestranking list
* Examine in PyMol
* Compare to wet lab results (enzyme assay)

In your lab notebook, record your steps and which library you have screened.

**PROTEIN PREPARATION**

When protein files are obtained from the Protein Data Bank, they are in more of a raw format. The PDB file needs to be cleaned up before a virtual screening run can be done on it. We will add hydrogens to those atoms which need them so that hydrogen bond interactions can be more accurately assessed. Waters molecules will also be removed to speed up docking- unless we know of a certain water molecule that should stay. Lastly, the active site will be defined by using whichever ligand is currently in the structure from the X-ray crystallography experiment. Then this ligand will be extracted so that we can dock the new compounds into the active site.

We will use the visual interface (GUI – graphical user interface) of GOLD along with its companion program **Hermes** to carry out these preparation steps. We will walk through the protein set up check list and concurrently generate a **gold.conf** file which determines the parameters for our docking.

**DOWNLOAD THE CRYSTAL STRUCTURE**:

For this lab, you will get to pick which crystal structure you would like to use. To select yours, g**o to the PDB website and search for entries of ‘human PTP1b’**. There are about 110 human PTP1b hits in the PDB

Select one that satisfies the following criteria:

* + Has only 1 inhibitor present in the active site (cannot be allosteric either)
  + Choose a file that has no mutations in the amino acid sequence
  + Select one with a Resolution for the X-ray structure that is < 2.0 Angstroms [Å]
  + Single chain – i.e. 1 polymer
  + Should not be a truncated protein (Number of residues ~ 298)

On the rhs (right hand side) of the PDB page, there is a link to the text file from the ‘right click’ menu.

Download the file to your Desktop (or wherever you want)

**Login to the DDFE using WinSCP**

ddfe.cm.utexas.edu (UTEID for both login and pass)

In virtual screening, it is important to keep your file structure organized and to reduce redundant files.

Create a directory structure on the DDFE like this below. You will be making 1 new directory for now.

e.g. /home/chem204/2011/YOURUTEID/LabVS3PTP1b

Transfer the text file of your protein (.pdb) from your local computer over to your directory in the DDFE.

**Connecting to the graphical interface for GOLD**

Make remote connection to DDFE using a graphical user interface (GUI) for GOLD

Open Xming server

Go to Start, Programs, Xming, Xming

Open Xlaunch

Go to Start, Programs, Xming, XLaunch

Select ‘Multiple Windows’

Select ‘Start no client’,

Skip next screen by selecting ‘Next’ then ‘Finish’ on next screen

Open **Putty** in Programs

Connect to Host Name: ddfe.cm.utexas.edu on Port 22 using SSH

On the left side of the window, Select the ‘SSH’ tab and then the ‘X11’ or ‘Tunnels’ tab

‘Enable X11 forwarding’

X display location: leave blank or enter localhost:0

# this is the default display on your computer

‘Open’

Login as user: type your user name for the DDFE (your UTEID)

Enter password

Type ‘ls’ to see the contents and ‘cd’ to change directories

Due to license issues, we will be using Gold 5.0 to setup our protein and then Gold 4.1 to run the screening jobs. Move your terminal window in to the folder where the .pdb file is for your protein. Then use the following command to force GOLD to use version 5.0

**$setgold50**

Then type this to open gold with the graphical user interface:

**$gold**

Ignore the ‘BadFont’ error message, if present

Don’t load a Conf file at the top (that is what you will be making here)

Step through the Configuration Options to set up your file

Skip Wizard

Skip Templates

**Protein > Load protein** “nameofyourpdbfile.pdb”

Gold 5.0 has a separate window for Global Options and a specific window for operations on your protein.

**Under Global Options:**

Define Binding Site –‘Select One or more ligands’

‘One or more ligands’ - choose the single ligand

‘Select all atoms within 7.5 Angstroms

Leave ‘Generate a cavity’ unchecked

Check – ‘Detect cavity’

Check – ‘Force all H bond donors/acceptors ….”

– verify active site in image on the Hermes visualizer

(only a small region around the ligand of the protein will be highlighted in gray)

**Under the tab for your PDB file name (to the right):**

Protonation & Tautomers > Add Hydrogens

For your report,Write down how many hydrogens added.

Extract/Delete Waters: Delete Remaining Waters (don’t select any of them to save).

For your report,Write down how many waters removed.

**Delete Ligands**

If there is more than one ligand,

you will need to go into the Hermes visualizer window to figure out which ligand

Go to View >> Protein Explorer

Click on the ‘+’ (plus sign) to see the different objects.

Extract, save as ‘LigandExtracted.mol2’

(this will be saved for defining the cavity site)

Side Note: Saving this extracted ligand can also allow you to re-dock it as a validation dock to compare the GOLD pose to the Xray crystallography pose and determine the RMSD value between the two.

**Back in WinSCP** – make sure your LigandExtracted has an extension

If not, then add it to the file (just add **.mol2** to the end)

Skip the remaining options for the protein.

In the Gold GUI – go back to **Global Options**

Select Ligands – you will need to go up in the directory tree until you get to the **/chem204/DatabasesVDS/LabVS3\_PTP1bLibrary** directory.

Then find the ligand library that you want to screen:

e.g. CB5k\_?all.sdf

This is the file you need to link to for your ligand library.

Then make sure the number of conformations per ligand or **GA Runs** is set to ‘10’

Skip the Reference Ligand

Skip ‘Configure Waters’

Skip ‘Ligand Flexibility’

Leave the defaults for ‘Fitness & Search Options’

‘GA Settings’ – 10%

Output Options – Change Output directory to ‘**ResultsCB5kRun1**’

UNCHECK – save ligand rank (.rnk) files

UNCHECK – save ligand log files

UNCHECK – save initialized ligand files

Save solutions to one file:

‘YourTargetvsYourLibraryRun1.sdf’ e.g. “PTP1bvsCB5kRun1.sdf”

bestranking\_list\_filename

‘BestYourTargetvsYourLibraryRun1.lst’ e.g. “BestPTP1bvsCB5kRun1.lst”

Skip ‘Information in File’

Under the ‘Selecting Solutions’ tab – select

‘Keep the top-ranked solutions for the best ???? ligands only’

This should be 10% of total

Follow instructions in DatabasesVDS folder on how to count the number of ligands (if you didn’t before)

Skip GoldMine

Skip Parallel GOLD – we will run in parallel but it will be executed remotely instead of at this console

Skip ‘Constraints’

‘Atom Typing’ - **Automatically set atom and bond types (for the ligand only)**:

Make sure only one box is checked - ‘Ligand’ only

At the top of the page hit Save

Hit ‘Finish’ to save the file

Save GOLD conf file as gold.conf

Save protein as PDBname\_protein.mol2

Then close GOLD/Hermes

**Now go back to WinSCP to VERIFY your newly made gold.conf file and MODIFY it for GOLD 4.1**

Set Autoscale to 0.1

cavity\_file = Cavity file name that you made in the Hermes prep - YourLigand.mol2

-may need to add the extension manually if not present

(do this on the actual ligand file and on the line in gold.conf file)

ligand\_data\_file = Reference to ligand file set (the whole path needs to be there).

Number of conformers is ‘10’

Be sure that ‘set\_protein\_atom\_types = 0’

directory = CB5kRun1

REMOVE this line - this is a Gold 5.0 command only.

match\_ring\_templates = 0

MODIFY this line- change a ‘1’ to a ‘0’

solvate\_all = 0

concatenated\_output = verify it matches what you entered before

clean\_up\_option save\_best\_ligands = ??? this is 10% of total

MODIFY this line- change a ‘1’ to a ‘0’

relative\_ligand\_energy = 0

protein\_datafile = Protein target file name (PDBname\_protein.mol2 – from Hermes prep)

Obtain an old hosts file from one of your previous runs and modify to do:

Gold.hosts 🡪 a total of **4** processes

Only use the upper half of the cluster (i.e. blades 8-15). Pick two different blades.

For Example: Where ‘X’ and ‘Y’ are the different blade numbers

compute-0-X.local 2

compute-0-Y.local 2

no\_of\_processes 4

**Logout of WinSCP and then log back in !**

Verify your files with Dr. B or a mentor before running!

BLADE RULES:

You are only allowed to run ONE job at a time.

You are only allowed to use 6 processors unless given permission to use more.

We must use GOLD version 4.1 to run the docking jobs

**$setgold41**

Run preliminary gold job using the proper command (Replace the ?? with the number of processors).

**$goldremoteP ?? gold.conf &**

If it does not run, Save the gold.err file and show to a Mentor or Dr. B

After the Run1 completes.

Verify the bestranking.lst has ~500 ligands

Verify poses by transferring the output SDF file to the Desktop and opening in PyMol. Are they all there and do they look ok?

Make note of how long it takes to run by looking at the ‘Date Created’ info for the files

**Run secondary gold job**

Copy the **gold.conf** and call it **gold2ndRun.conf**

Make changes to the **gold2ndRun.conf**

Change the Autoscale to 1.

Change your input ligand file so that it is the ‘**concatenated\_output’** from your first run

(put the link to the file that is now in the newly made Results folder)

Output Options – Change Output directory to ‘**ResultsCB5kRun2**’

Change Bestranking to “BestPTP1bvsCB5kRun2.lst”

concatenated\_output = “PTP1bvsCB5kRun2.sdf”

For the ligand solutions, save 10% of what you put in from the first run.

clean\_up\_option save\_best\_ligands 50

Start Run2 with the command (Replace the ?? with the number of processors). Be sure you use gold2.conf

**$goldremoteP ??gold2ndRun.conf &**

Make note of how long it takes to run by looking at the ‘Date Created’ info for the files

**Create Excel spreadsheet** – use the text import wizard

Sort the ligands by GOLD score using the bestranking.lst file

Number the top ones

* Do the 5 known binders from BindingDB show up in the 2nd Run list? If so, where? If not, where do they show up in the 1st Run List.
* Did the 4 VDS compounds from the Fall show up in the 2nd Run list? If so, where? If not, where do they show up in the 1st Run List.
* Where does the one untested compound show up? (5648649 – Chembridge)

**For Lab Report:**

See **Lab Enzyme Inhibition Assay** for general guidelines on this combined final report.

**For your Bestranking Table in the Lab Report**

Show the Bestranking.lst table in Excel for the top 50 inhibitors from the 2nd GOLD run.

If the 5 known inhibitors form BindingDB aren’t in there – add them.

If the 4 VDS compounds are not there – add them

If the untested compound (5648649) is not there – add it

Your table should have at least 50 compounds and then a maximum of 50+5+4+1 = 60 compounds

If these other compounds aren’t in the top 50 – you will need to find them in the top 500 from your 1st Run.

Then you can list their rank as simply > 50

**Show PyMol Images and Lipinski’s Info ONLY for the following ligands:**

* 1. The best compound from Chembridge (i.e. the CB5k library)
  2. The best known inhibitor from BindingDB
  3. The compound that you tested in the wet lab

Lipinski’s: To find the information, go to these sites:

**Chembridge** – <https://www.hit2lead.com/> Go to Screening Compounds, Search by ID

**Maybridge** – <http://www.maybridge.com/>

Login as [VDSclass@gmail.com](mailto:VDSclass@gmail.com) Password is painter214

Go to Screening Compounds, Use the Cat\_No in the SDF file to search for the ligand.

Remove the final letters. e.g. KM10410SC needs to be changed to ---> KM10410

* + Also Ryan Scientific has info on Maybridge ligands

**BindingDB** - <http://www.bindingdb.org> See the hyperlinks at the

**BindingDB\_50243240**

<http://www.bindingdb.org/bind/chemsearch/marvin/MolStructure.jsp?monomerid=50243240>

**BindingDB\_50131107**

<http://www.bindingdb.org/bind/chemsearch/marvin/MolStructure.jsp?monomerid=50131107>

**BindingDB\_50170986**

<http://www.bindingdb.org/bind/chemsearch/marvin/MolStructure.jsp?monomerid=50170986>

**BindingDB\_50131106**

<http://www.bindingdb.org/data/mols/tenK5013/MolStructure_50131106.html>

**BindingDB\_50228029**

<http://www.bindingdb.org/bind/chemsearch/marvin/MolStructure.jsp?monomerid=50228029>

To find the Molecular Properties (Lipinski’s Rule of Five info) for these ligands, click on the ‘**Ligand Links**’.

PyMol Images:

Show the docking poses in the active site of PTP1b in a similar fashion as the VDS2 lab.